

AINA

AI and Analytics

Volume 3.0 Edition 2021-2022



Public Policy
Metaverse
Blockchain

Astronomy
Jurisdiction
Language Model

Data Driven Governance

An intersection of Data Science and Public Policy

Shreya Mehrotra



Public policy comprising of national constitutional laws, rules, rulings from courts, and local ordinance is truly essential to meet the needs of citizens and protect society's values as a whole. Until recently, it was believed to be just a field of study meant for government officials involved in the decision-making process that affects the public at large. But with rise of journalism and media, the interest and awareness through debate and discussion about the policy changes has increased manifold among the masses.

With the rise in the era of Internet, both companies and individuals have been overwhelmed with the amount of data and information they are constantly being exposed to. Companies have witnessed several waves of data, which includes e-mail, web traffic, comments from social networks, and ever-sophisticated computer tracking of shipments, sales, suppliers, and customers. This has led to a rapid increase in the volume of data and the application of data science in almost every aspect of the corporate world, but there still lies an untapped area of public policymaking where the scope of analytics can be further expanded. Information ought to be the most valuable resource in a modern economy since it can be used to create new goods and services, make wiser decisions, provide businesses a competitive edge, and increase growth and productivity. In response to such rising trends, governments today around the world endeavour to utilize data science and big data to develop public policy. How and in which areas data can be used remains a topic for research and debate.

Analytics in Election Campaigning

The term “data scientist” was hardly known to the top-tier policymakers till the early 2010s. It was only after the 2012 US presidential elections that many researchers and academicians studying policy started to discover that a substantial amount of data analytics was used by

“There’s a digital revolution taking place both in and out of government in favor of open-sourced data, innovation, and collaboration”

Barack Obama to gain a competitive edge in the elections. This was the first instance when individual voter data was used to predict and forecast the results of the election. The primary indicator of concern was the ratio of the number of people who were likely to vote for Obama to the total number of people voting. It was identified that there were three main aspects that had to be catered to : registration, persuasion and turnout. Special care was taken to encourage the uninterested voters to enhance the turnout, and along with that, the undecided voters were targeted during the election campaigns. This approach had indeed proved to be a game changer for Obama’s administration in not only winning the presidential elections but also advocating data as a new currency for knowing the citizens better.

Obama’s administration has further served as an example which has influenced law makers to think in a direction to use data science in existing policy making. Henceforth, election campaigns throughout the world, be it Indian Lok Sabha elections, US presidential elections have started using data for targeted campaigning to attract potential voters.



Earlier, the concept of segmenting customers for advertising and selling products was heavily used in the market, but this instance of segmenting citizens and identifying potential voters has played a pivotal role in Obama's victory hence outlining the potential use of data for making policy decisions.

This concept of exploiting data driven campaigning has marked the beginning of a new way of thinking where the policy makers can segment citizens of their country in order to formulate policies. Governments around the world have attempted to figure out best ways of transforming this approach efficiently in which the voters can be seen as citizens and the election campaigns can be extended to policies. This has been coupled with the continued increase in data generation and availability with the excessive use of internet which is being used in many forms in deriving policies. Data is generated with every click, where the behaviour of every user is recorded in the form of digital time stamps. Time stamps are generated for multiple users during different times of the day and stored for a short span.

This data can further be used to forecast and derive conclusions about population preferences that can be further used for policymaking. But alas, there remains a gap between theory and practice. The actual use of data by policymakers to formulate meaningful policy decisions still remains a remote concept. Optimistically, the silver lining in all these discussions is the emergence of research trends in order to build a concrete understanding of the behaviours and preferences of the masses and predict social phenomena. This research is further extended to be adopted in the public realm during times of crisis such as war. Such a crisis creates a situation of extreme environment that favors the formulation of powerful technical innovations enabling the concerned officials to rely on them to take decisions.

The COVID-19 pandemic has been nothing less than a war. It has proved to be a perfect opportunity for governments and other civic bodies to realize their individual and joint potential in harnessing innovative technologies to make decisions for the betterment of the masses. However, there still lie a number of challenges in the practical implementation of these technologies. One of the primary matters of concern has been Data Privacy.



"In holding scientific research and discovery in respect, as we should, we must also be alert to the equal and opposite danger that public policy could itself become the captive of a scientific-technological elite"

President Dwight Eisenhower

Data Privacy- A concern

Quite a few governments attempted to control the spread of the coronavirus via contact tracing, like the COVID Safe app by the Australian government, Corona 100 started in Korea and Aarogya Setu in India. On similar lines, applications were developed by the Singapore government and their use was mandated.

The Singapore government has developed two separate applications for contact tracing: SafeEntry and TraceTogether. SafeEntry is a logging system that is required while scanning codes that allows users to check in to public venues. TraceTogether is a Bluetooth technology that is used to track whether users have come into close contact with people infected with COVID-19. The main difference between both the applications is that users need to provide their details like contact numbers and identification IDs on SafeEntry, whereas, TraceTogether assigns random temporary IDs. Both the applications are interlinked so that the data from TraceTogether can be used while scanning data from SafeEntry. The individuals infected with COVID-19 are obligated to provide

their TraceTogether data to the Ministry of Health. This ensures the privacy of individual data as the precise location history is anonymized, and personal information such as mobile numbers are encrypted. The SafeEntry data available with the government authorities is stored locally and is also deleted after 25 days.

Despite such a design that kept in mind the privacy concerns of individuals, citizens were highly skeptical about providing their personal data and how it was being used. This concern was further fuelled by the revelation that the data collected by TraceTogether was used by police for the investigation of severe crimes such as murders. This has posed a pressing need for the formulations of methods of data collection, keeping in mind the privacy concerns of citizens. Efforts to be transparent to the public should also be made more carefully to minimize room for misunderstanding and backtracking. Public communication and inter agency coordination will be increasingly important as emerging technologies become more frequently designed and adopted.

The COVID-19 pandemic served as a window of opportunity to accelerate the development and adoption of innovative technologies. Applications for contact tracing have emerged as a creative and effective replacement for conventional contact tracing, and some governments have used these applications in order to take decisions as to which areas to open up after lockdown while staying on guard against COVID-19. However, the hasty deployment and lack of testing time of these applications has led to an increased skepticism about the use of public data even after the pandemic. The government needs to be more transparent about the purging of public data.

While opting for a data centric approach and regulating long-lived data technology, governments must carefully design their applications taking care of these privacy issues to prevent unintended and perhaps disastrous side-consequences.



Data Analytics in Judiciary

The power of data analytics as evident from above examples can be used to collect data about the citizens. There are various other areas in which the use of data science can be leveraged that can be used by the government in better managing their citizens. One such area where analytics is currently being used is the judiciary. The upcoming lawyers and legal practitioners are increasingly enthusiastic about using data to automate their manual day-to-day tasks.

Deep learning technologies like Natural Language Processing and Cognitive Computing are being used to mine an enormous number of legal documents. The government is using more advanced technological tools for fraud detection. A new initiative, 'Project Insight', has been laid by the government to catch hold of tax evaders by gathering the income status of people from their social media accounts. This would save a massive amount of government resources and budget that it currently spends on monitoring illegal activities during ITR Filings.

Data Collection

Despite the use of data by the government in various small-scale instances, there remains a potential for wide-scale adoption of data in making policy decisions. One of the leading challenges that come into the picture is the collection of secured quality data. Most of the data generated these days comes from unsolicited sources of information, making it extremely difficult to filter out meaningful ones. Hence, it is imperative for the policymakers to spend a good amount of resources in hiring the right people, known as 'Data Scientists'. These people can also come from third-party consulting firms.

One such example is the city-level governments in US that have expanded the scope of their analytics teams to include more data science capabilities. One of the potential challenges that might come up in this process is the capture of real-time data to understand the changing preferences and behavior of the masses. Several civic tech communities like Chi Hack Night have come up to cater to this issue, using the accessible open-source government data in their projects. The analysis and findings can also serve as sources or pipelines for further data analysis.





In the picture :Carving at the Lincoln memorial. As governments start using analytics for policy making, they must not forget how the policy affects every single stakeholder. Lincoln has rightly said a government is of the people , by the people and for the people

In a country like India, having a population of around 135 crores, the need for analytics arises because of the heavy amount of investment in the collection of data during the census. An excellent solution to the above problem can be the use of AI and computer vision in order to gather data. A significant amount of research has been ongoing on the use of AI surveillance technologies in order to track and monitor citizens without violating human rights. At least seventy-five out of 176 countries globally are actively using AI technologies. This includes facial recognition systems (sixty-four countries) , smart policing (fifty-two countries) and smart city/safe city platforms (fifty-six countries). Such platforms prove to be extremely beneficial in capturing data for the real-time modelling and analysis.

Human data is captured using satellite images, and the observations derived from them can be used to formulate lifestyle patterns such as clustering areas of high or low densities. Density can be used to classify people on the basis of their income and hence can be used to identify areas of low or high-income groups. This can further be extrapolated and utilized in predicting the poverty levels of the masses. Researchers have conducted studies that have shown that AI surveillance can be used to curb military expenditures. These technologies can be used in the defense sector by governments in order to detect unsolicited activities or provide prompt aid in times of emergency and crisis situations. Such research patterns still remain in their infancy stages but can still be seen as stepping stones towards a technology-driven economy.

